ESS Update

M. Nussbaum

PINSKER BOUND

COMMUNICATION THEORY, STATISTICAL

DENSITY ESTIMATION

FISHER INFORMATION

KERNEL ESTIMATORS

LINEAR ESTIMATORS, BAYES

LOCAL ASYMPTOTIC NORMALITY

METHOD OF SIEVES

MINIMAX ESTIMATION

NOISE (SIGNAL PROCESSING IN THE PRESENCE OF)

PREDICTION AND FILTERING; LINEAR

SIEVES, METHOD OF

SPECTRAL ANALYSIS

SHRINKAGE ESTIMATORS

SMOOTHNESS PRIORS

SOBOLEV SPACES

SPLINE FUNCTIONS

STATIONARY PROCESSES

STEIN EFFECT

The Pinsker bound describes the exact asymptotics of the minimax risk in a class of nonparametric smoothing problems where root-n consistent estimators do not exist. The result from 1980 (Pinsker [46]) represents a breakthrough in nonparametric estimation theory, by allowing comparison of estimators on the level of constants rather than just comparing rates of convergence. For the minimax risk, the Pinsker bound provides not only the optimal rate of convergence for estimators, but also the "optimal constant". Such optimal constants are well known for estimation in regular parametric models, and usually given by the asymptotic Fisher information. The Pinsker bound can be established in a variety of problems (density estimation, nonparametric regression, signal estimation in Gaussian white noise, spectral density estimation of a stationary Gaussian process and others). But the result is closely connected to special loss functions and a priori smoothness classes, essentially to a Hilbert space setting.

OPTIMAL RATES AND OPTIMAL CONSTANTS

Consider estimation of a probability density f from independent, identically distributed random variables $X_1, \ldots, X_n$, and assume that $f \in \mathcal{F}$- a class of smooth functions on the unit interval. Let $\hat{f}_n$ be an estimator and consider the integrated mean squared error $\left\|\hat{f}_n - f\right\|_2^2$, where $\|\cdot\|_2$ is the norm in the Hilbert space $L^2(0,1)$. An estimator $\hat{f}_n$ is said to attain an optimal rate of convergence $r_n \to 0$ if for some constant c_1

$$\sup_{f \in \mathcal{F}} E_{n,f} \left\|\hat{f}_n - f\right\|_2^2 \leq r_n^2 \; c_1(1 + o(1)), \tag{1}$$

and no estimator can attain a better rate: for a $c_2 > 0$

$$\inf_{\hat{f}_n} \sup_{f \in \mathcal{F}} E_{n,f} \left\|\hat{f}_n - f\right\|_2^2 \geq r_n^2 \; c_2(1 + o(1)) \tag{2}$$

where the infimum is taken over all estimators. A shorthand notation is

$$R_n(\mathcal{F}) := \inf_{\hat{f}_n} \sup_{f \in \mathcal{F}} E_{n,f} \left\|\hat{f}_n - f\right\|_2^2 \asymp r_n^2 \tag{3}$$

where $a_n \asymp b_n$ for sequences means that there are finite positive constants c_1, c_2 such that $c_2 + o(1) \leq a_n/b_n \leq c_1 + o(1)$. In the basic nonparametric cases $r_n = n^{-m/(2m+1)}$ where m is the degree of smoothness of functions in $\mathcal{F}$. For instance, $\mathcal{F}$ might be the class of densities on $[0,1]$ such that for a given constant M, the m-th derivative exists and is everywhere bounded by M. There is a large variety of such results in density estimation and related nonparametric smoothing problems, mostly in the context of the method of sieves*. Rate optimality is a natural first concept of asymptotic efficiency when estimators with $r_n = n^{-1/2}$ (root-n-consistent estimators) do not exist and Fishers bound for asymptotic variances does not apply. But this concept is unsatisfactory in a sense: the constants c_1 and c_2 are not specified; they are only required to be positive. Thus no matter how large c_1 in (1) is compared to c_2 in (2), the estimator is still deemed asymptotically optimal.

For nonparametric estimation problems where the optimal rate is slower than $n^{-1/2}$ it seemed remote for a long time that coinciding constants c_1 and c_2 might be found, and a thus a sharper optimality criterion than (1) and (2) be made available. In fact results like (3) were first established as pure existence theorems for constants c_1, c_2 (cp. Ibragimov and Khasminski [37], chap. 4) and it is frequently hard to get quantitative information on them. The Pinsker bound achieves just that, by finding "exact constants" $c_1 = c_2$ for certain functional classes $\mathcal{F}$.

Pinsker bound for density estimation. It is essential that $\mathcal{F}$ is a Sobolev class or a function set with similar structure. For given $M > 0$ and natural m, a Sobolev class $W^{m,2}(M)$ of functions f on $(0,1)$ consists of $m-1$ times differentiable f; the derivative $f^{(m-1)}$ is required to be the Lebesgue integral of some function $D^m f \in L^2(0,1)$, and $\|D^m f\|_2^2 \leq M$. (Here $f^{(0)}$ is taken to be f; for given m, the union of these classes over $M > 0$ is the Sobolev space* $W^{m,2}(0,1)$). Assume that $\mathcal{F}$ is given by all densities in

$W^{m,2}(M)$. Let $R_n(\mathcal{F})$ be defined by (3) and write $a_n \sim b_n$ for sequences if $a_n/b_n = 1 + o(1)$. Then

$$R_n(\mathcal{F}) \sim r_n^2 \, (M/\pi^{2m})^{1/(2m+1)} \, P_m, \; n \to \infty \tag{4}$$

where $r_n = n^{-m/(2m+1)}$ and

$$P_m = \left(\frac{m}{(m+1)} \right)^{2m/(2m+1)} (2m+1)^{1/(2m+1)} \tag{5}$$

is the PINSKER CONSTANT. This result for density estimation is essentially due to Efromovich and Pinsker [8], which built upon the basic paper [46]. The optimal rate $R_n(\mathcal{F}) \asymp n^{-2m/(2m+1)}$ was known before in density* estimation and similar problems, cf. the survey paper of Ibragimov and Khasminski [35].

Analogy to Fisher information bound. The Pinsker bound can be compared to Fisher's bound for asymptotic variances in regular parametric models. Suppose the density f is in a parametric family $(f_\theta, \theta \in \Theta)$ where $\Theta \subset R^k$ is open and bounded, and the family has finite nonsingular Fisher information matrix $I_F(\theta)$ for every $\theta \in \Theta$. Then (with more regularity and moment assumptions) there are estimators $\hat{\theta}_n$ attaining Fisher's bound, and this bound cannot be improved. A modern (local asymptotic minimax) formulation for this is: there is an estimator $\hat{\theta}_n$ such that for every open set $A \subset \Theta$

$$\sup_{\theta \in A} E_{n,f} \left\| \hat{\theta}_n - \theta \right\|^2 \leq n^{-1} \sup_{\theta \in A} \mathrm{tr}[I_F^{-1}(\theta)](1 + o(1)) \tag{6}$$

and this bound cannot be improved:

$$\inf_{\hat{\theta}_n} \sup_{\theta \in A} E_{n,f} \left\| \hat{\theta}_n - \theta \right\|^2 \geq n^{-1} \sup_{\theta \in A} \mathrm{tr}[I_F^{-1}(\theta)](1 + o(1)). \tag{7}$$

(cf. Ibragimov and Khasminski **??**.) Here $\mathrm{tr}[\cdot]$ is the trace of a matrix and $\|\cdot\|$ is Euclidean norm. For the case $A = \Theta$ this means

$$R_n(\Theta) := \inf_{\hat{\theta}_n} \sup_{\theta \in \Theta} E_{n,f} \left\| \hat{\theta}_n - \theta \right\|^2 \sim n^{-1} \sup_{\theta \in \Theta} \mathrm{tr}[I_F^{-1}(\theta)]. \tag{8}$$

where $a_n \sim b_n$ means $a_n = b_n(1 + o(1))$. So both (8) and (4) are improvements of (1), (2), in the sense that the constant c_2 is specified, and an estimator can be found such that $c_1 = c_2$. In (8) the problem is parametric (smoothly indexed by a k-dimensional parameter $\theta \in \Theta$) and the optimal rate is $n^{-1/2}$, while in the Pinsker bound (4) the parameter set $\mathcal{F}$ is infinite dimensional with a slower optimal rate. In this sense the Pinsker bound is a nonparametric analog of the Fisher information bound.

ESTIMATING A BOUNDED NORMAL MEAN

The connection to finite dimensional problems can be illustrated in a very simple Gaussian model. Suppose we observe a k-dimensional Gaussian vector

$$Y = \theta + n^{-1/2}\,\xi, \tag{9}$$

where ξ is standard Gaussian in R^k and the problem is to estimate the k-dimensional parameter θ with squared Euclidean loss $\|\cdot\|^2$. The parameter space is a ball in R^k: $\Theta = \{\theta : \|\theta\|^2 \leq M\}$. Let as in (8) $R_n(\Theta)$ be the minimax risk over all estimators $\hat{\theta}_n$ with squared Euclidean loss.

Parametric information bound. Suppose first that k is fixed and $n \to \infty$. This is a finite dimensional parametric model, of the very regular type assumed with (6), (7) (except for the inessential difference that Θ was assumed open there). Indeed we can construe Y in (9) as a sufficient statistic (sample mean) from i. i. d. observed Gaussian vectors Y_i with expectation θ and unit covariance matrix. The Fisher information matrix $I_F(\theta)$ in (8) then is the unit matrix, for all $\theta \in \Theta$, and (8) takes the form

$$R_n(\Theta) \sim kn^{-1}, n \to \infty. \tag{10}$$

Pinsker bound. Suppose now that k increases with n: $k/n \to K > 0$. It can then be shown that

$$R_n(\Theta) \sim \frac{MK}{M+K},\ n \to \infty. \tag{11}$$

Note that the size M of the ball now appears in the risk asymptotics, contrary to (10).

Attainment of the bounds. The upper bound parts of (10) and (11) are very easy to establish here. Let $c \in [0,1]$ be a real number; for the shrinkage estimator* $\hat{\theta}_n^c = cY$ one has a bias-variance decomposition

$$E_{n,\theta} \left\| \hat{\theta}_n^c - \theta \right\|^2 = (1-c)^2 \left\|\theta\right\|^2 + c^2 kn^{-1} \leq (1-c)^2 M + c^2 kn^{-1}.$$

Minimizing over c yields

$$R_n(\Theta) \leq Mkn^{-1}/(M + kn^{-1})$$

with an optimal $c = M/(M + kn^{-1})$. In the parametric case when k is fixed, $c \to 1$ and the upper bound part of (10) results. In the Pinsker case, where $kn^{-1} \to K > 0$, the bound of (11) is attained.

In this simple model the Pinsker bound is obtained as the result of a dimension asymptotics effect when estimating a bounded normal mean in Euclidean space. The conceptual link between (4) and (8) becomes apparent. The connection with shrinkage and the Stein effect* is further discussed by Beran [3]

SIGNAL ESTIMATION IN GAUSSIAN WHITE NOISE

Let us formulate the Pinsker bound in the so-called Gaussian white noise model. This is a continuous version of (9), which is also called sometimes "continuous (nonparametric) regression". Consider an observed Gaussian stochastic process

$$Y(t) = \int_0^t f(u)du + n^{-1/2}\ W(t),\ t \in [0,1] \tag{12}$$

where $W(t)$ is standard Brownian motion and $n \to \infty$. When the function f is in $L^2(0,1)$, this can equivalently be written in stochastic differential equation* form

$$dY(t) = f(t)dt + n^{-1/2}dW(t),\ t \in [0,1],\ Y(0) = 0 \tag{13}$$

where $dW(t)$ is the derivative of $W(t)$,i. e. Gaussian white noise. (The boundary condition $Y(0) = 0$ can be suppressed for the statistical equivalence). This model occurs in communication theory*; it was recognized by Ibragimov and Khasminski [37] as being of great theoretical value in mathematical statistics. The process $Y(t)$ is a diffusion process* with drift $F(t) = \int_0^t f(u)du$. The function f, called drift density or "signal", turns out to be an analog of the probability density in the case of independent identically distributed random variables, as far as statistical inference is concerned. It is instructive to formulate and study estimation and testing problems under assumptions $f \in \mathcal{F}$, both parametric and nonparametric. Since observations are exactly Gaussian (not just asymptotically normal in distribution) and f need neither be positive nor integrate to one, the model can serve as an idealized version of many other statistical problems. The Gaussian white noise model (also called the signal recovery model) has thus become a prime object of study in asymptotic statistics, especially in nonparametric settings. Pinsker's result [46] was first established in the Gaussian white noise model, thus confirming its pivotal role.

Actually the result was developed in a discrete version of (13). Take an orthonormal basis of $L^2(0,1)$, $(\varphi_j)_{j=1}^{\infty}$ say, and consider observed numbers Y_j, given by stochastic integrals $Y_j = \int \varphi_j(t)dY(t),\ \ j = 1, 2, \ldots$. It is well known that these can be represented, for $\epsilon = n^{-1/2}$, as

$$Y_j = \theta_j + \epsilon\, \xi_j,\ j = 1, 2, \ldots \tag{14}$$

where $\theta_j = \theta_j(f)$ are the Fourier coefficients of f in the basis $(\varphi_j)_{j=1}^{\infty}$ and ξ_j are independent standard Gaussian random variables (in fact $\xi_j = \int \varphi_j(t)dW(t)$). The process $Y(t)$ can then be reconstructed from the Y_j, and so the models (13) and (14) are equivalent in a statistical sense if $\theta_j = \theta_j(f)$.

The sequence $\theta = (\theta_j)$ is in the space l^2 (the space of square summable sequences), which is isomorphic as a Hilbert space to $L^2(0,1)$. A central assumption for the Pinsker bound is that the function set $\mathcal{F}$ can be represented as an ELLIPSOID. An ellipsoid in l^2 is a set

$$\Theta = \left\{ \theta \in l^2 : \sum_{j=1}^{\infty} a_j \theta_j^2 \leq M \right\}. \tag{15}$$

for certain nonnegative numbers $(a_j)_{j=1}^{\infty}$ and M. In the discrete model (14), consider the problem of estimating the sequence θ with a loss given by $\|\cdot\|_{l^2}^2$, the squared norm in l^2 (i. e. $\|\theta\|_{l^2}^2 = \sum_{j=1}^{\infty} \theta_j^2$), under an assumption $a_j \to \infty$ as $j \to \infty$. A LINEAR FILTER is a sequence $c = (c_j) \in l^2$ such that $0 \leq c_j \leq 1$ for all j. For such a c, a LINEAR FILTERING ESTIMATE of θ is given by $\hat{\theta}^c = (c_j y_j)$. Consider the minimax estimator within this class: define

$$R_{L,\epsilon}(\Theta) = \inf_c \sup_{\theta \in \Theta} E_{\epsilon,\theta} \left\| \hat{\theta}^c - \theta \right\|_{l^2}^2. \tag{16}$$

Along with the minimax risk over this restricted class of estimators, consider the risk over arbitrary estimators $\hat{\theta}_\epsilon$ (analogous to (3)):

$$R_\epsilon(\Theta) = \inf_{\hat{\theta}_\epsilon} \sup_{\theta \in \Theta} E_{\epsilon,\theta} \left\| \hat{\theta}_\epsilon - \theta \right\|_{l^2}^2. \tag{17}$$

In this framework, Pinker's result takes the following remarkable form: if $R_{L,\epsilon}(\Theta)/\epsilon^2 \to \infty$ then

$$R_\epsilon(\Theta) \sim R_{L,\epsilon}(\Theta), \ \epsilon \to 0. \tag{18}$$

In words, the minimax linear filtering estimate is asymptotically minimax among all estimators. The asymptotics of $R_{L,\epsilon}(\Theta)$ can often be found as regards rates and constants, and then gives rise to results like (11) and (4).

Evaluating the minimax linear risk. The minimax linear filter is easy to calculate in the above framework, cf. Belitser and Levit [1]. The functional

$$L_\epsilon(c, \theta) = E_{n,\theta} \left\| \hat{\theta}^c - \theta \right\|_{l^2}^2 \tag{19}$$

has a saddle point $(c_\epsilon^*, \theta_\epsilon^*)$, so that

$$R_{L,\epsilon}(\Theta) = \inf_c \sup_{\theta \in \Theta} L_\epsilon(c, \theta) = \sup_{\theta \in \Theta} \inf_c L_\epsilon(c, \theta) = L_\epsilon(c_\epsilon^*, \theta_\epsilon^*). \tag{20}$$

This saddle point can be found explicitly; the optimal estimator is then given by

$$c_{\epsilon,j}^* = (1 - (\mu_\epsilon a_j)^{1/2})_+, \; j = 1, 2, \ldots \tag{21}$$

where $x_+ = \max(x, 0)$ and μ_ϵ is the unique solution of a certain equation. It remains to calculate the asymptotics of $R_{L,\epsilon}(\Theta) = L_\epsilon(c_\epsilon^*, \theta_\epsilon^*)$, as $\epsilon \to 0$, depending on a, M. The most important case is $a_j \sim (\pi j)^{2m}$, $j \to \infty$, where

$$R_{L,\epsilon}(\Theta) \sim \epsilon^{4m/(2m+1)} \; (M/\pi^{2m})^{1/(2m+1)} \; P_m \tag{22}$$

with P_m from (5). This coincides with (4) for $\epsilon = n^{-1/2}$. Here $\mu_\epsilon \to 0$ as $\epsilon \to 0$, so that $c_{\epsilon,j}^*$ in (21) exhibit the typical behaviour of smoothing or tapering coefficients (for fixed j each coefficient tends to 1, and the number of nonvanishing $c_{\epsilon,j}^*$ tends to infinity as $\epsilon \to 0$).

Sobolev classes as ellipsoids. The trigonometric orthonormal basis in $L^2(0,1)$ can be used to represent a Sobolev function class as an ellipsoid, when certain boundary conditions are added. Consider the periodic Sobolev class

$$\tilde{W}^{m,2}(M) = \left\{ f \in W^{m,2}(M) : \; f^{(k)}(0) = f^{(k)}(1), \; k = 1, \ldots, m-1 \right\}. \tag{23}$$

Let the trigonometric basis be $\varphi_1(t) \equiv 1$, $\varphi_{2k}(t) = 2^{1/2} \cos(2\pi k t)$, $\varphi_{2k+1}(t) = 2^{1/2} \sin(2\pi k t)$ for $k \geq 1$. Then $\tilde{W}^{m,2}(M)$ is an ellipsoid $\Theta = \Theta(a, M)$ for $a_1 = 0$, $a_{2k} = a_{2k+1} =$

$(2\pi k)^{2m}$, $k = 1, 2, \ldots$. The asymptotics of a_j is $a_j \sim (\pi j)^{2m}$ for $j \to \infty$, so that (22) obtains. The periodic Sobolev classes were the first function classes considered in the original result ([46]) and in the subsequent application to density estimation (Efromovich and Pinsker [8]).

For the classes $W^{m,2}(M)$ without boundary conditions, ellipsoid representations can be found using other bases $(\varphi_j)_{j=1}^{\infty}$; cf. the part on nonparametric regression below.

Renormalization and continuous minimax problem. Let us sketch a derivation of the asymptotics (22) by a renormalization technique (cf. Golubev [17]). Suppose that $a_j = (\pi j)^{2m}$ and consider linear filters $c_j = \phi(hj)$, where c: $[0, \infty) \mapsto [0, 1]$ is a "filter function" (assumed Riemann integrable) and h is a bandwidth parameter, tending to 0 for $\epsilon \to 0$. It can then be shown for the functional $L_\epsilon(c, \theta)$ from (19) that for a certain choice of h

$$\inf_c \sup_{\theta \in \Theta} L_\epsilon(c, \theta) \sim \epsilon^{4m/(2m+1)} (M/\pi^{2m})^{1/(2m+1)} \inf_\phi \sup_\sigma L_0(\phi, \sigma)$$

where

$$L_0(\phi, \sigma) = \int_0^\infty (1 - \phi(x))^2 \sigma^2(x) dx + \int_0^\infty \phi^2(x) dx.$$

and the supremum extends over continuous functions σ on $[0, \infty)$ fulfilling $\int x^{2m} \sigma^2(x) dx \leq 1$. The saddle point problem (20) is thus asymptotically expressed in terms of a fixed continuous problem, and the Pinsker constant P_m from (5) is the value of the game:

$$P_m = \inf_\phi \sup_{\int x^{2m} \sigma^2(x) dx \leq 1} L_0(\phi, \sigma).$$

The optimal function

$$\phi^*(x) = (1 - (\lambda^* x)^m)_+ \tag{24}$$

has sometimes been called the Pinsker filter (λ^* is a certain constant; cp. the form of c_ϵ^* in (21)).

The continuous saddle point problem arises naturally in a continuous Gaussian white noise setting (13) and a parameter space described in terms of the continuous Fourier transform (cf. Golubev [17]), e. g. a Sobolev class of functions on the whole real line. The Fourier transform of the filter $\phi^*(x)$ gives rise to a kernel estimator* attaining the Pinsker bound (Golubev [18], cf. also [40]).

BACKGROUND: BAYES-MINIMAX PROBLEMS

The term "optimal filtering" used by Pinsker in [46] points to a Bayesian aspect of the result, although it is the minimax risk which is evaluated. Consider the model (9) for dimension $k = 1$ and for $n = 1$: we observe a real Gaussian random variable

$$Y = \theta + \ \xi,$$

where ξ is standard Gaussian and the problem is to estimate θ with squared loss. The parameter space is an interval: $\Theta = \{\theta : \theta^2 \leq M\}$.

1. Minimax risk as least favorable Bayes. Let $R(\Theta)$ be the minimax risk and let $r(Q)$ be the Bayes risk for a prior distribution Q for θ, not necessarily concentrated on Θ. Denote supp(Q) the support of Q; it is well known in the general theory of minimax estimation* that

$$R(\Theta) = \sup_{Q:\,\mathrm{supp}(Q)\subset\Theta} r(Q). \tag{25}$$

2. Minimax linear risk. A linear estimator $\hat{\theta}^c$ is given by $\hat{\theta}^c = cY$ where c is a real number. Its risk is

$$E_\theta\,(\hat{\theta}^c - \theta)^2 = (1-c)^2\theta^2 + c^2 = L(c,\theta),$$

say. For given θ, the best linear estimator is given by $c(\theta^2) = \theta^2/(\theta^2+1)$, we have $0 \leq c(\theta^2) \leq 1$ (hence $c(\theta^2)$ is a linear filter), and the risk is $\theta^2/(\theta^2+1)$. In view of the minimax theorem (20), $\hat{\theta}^{c(M)}$ is minimax among linear estimators and the minimax linear risk is

$$R_L(\Theta) = M/(M+1).$$

3. Minimax linear risk as least favorable Bayes. For a prior distribution Q for θ having $E_Q\theta^2 = \sigma^2$, not necessarily concentrated on Θ, the integrated risk is again

$$E_Q E_\theta (\hat{\theta}^c - \theta)^2 = (1-c)^2 E_Q\theta^2 + c^2 = L(c,\sigma).$$

Hence $\hat{\theta}^{c(\sigma^2)}$ is also the Bayes linear* estimator for Q, with risk $\sigma^2/(\sigma^2+1)$. This estimator is Bayes among all estimators if Q is centered normal, i. e. $Q = N(0,\sigma^2)$. Hence

$$r(N(0,M)) = M/(M+1) = R_L(\Theta).$$

Moreover, Donoho and Johnstone [4] establish the following:

$$\sup_{E_Q\theta^2 \le M} r(Q) = r(N(0,M)). \tag{26}$$

Thus $R_L(\Theta)$ is also the solution of a Bayes-minimax problem:

$$R_L(\Theta) = \sup_{E_Q\theta^2 \le M} r(Q). \tag{27}$$

4. Bracketing the minimax risk. Since always $R(\Theta) \le R_L(\Theta)$, relations (25) and (27) imply the following bounds for the minimax risk:

$$\sup_{Q\colon \mathrm{supp}(Q)\subset\Theta} r(Q) \le R(\Theta) \le \sup_{E_Q\theta^2 \le M} r(Q).$$

In a k-dimensional model (9) for $n = k$, with parameter space $\Theta = \{\theta : \|\theta\|^2 \le M\}$ and squared Euclidean loss, we obtain analogously for the Bayes and minimax risks depending on n

$$\sup_{Q\colon \mathrm{supp}(Q)\subset\Theta} r_n(Q) \le R_n(\Theta) \le \sup_{E_Q\|\theta\|^2 \le M} r_n(Q).$$

This gives the basic heuristics for the validity of the Pinsker bound. By reasons of symmetry the set of Q in the upper bound can be restricted to product measures

$Q = Q_1^{\otimes n}$ with $E_{Q_1}\theta_1^2 \leq k^{-1}M$. These Q do not have support in Θ in general, but as $n \to \infty$ they tend to be concentrated on Θ as a law of large numbers effect, so that asymptotically the upper and lower brackets coincide.

The special role of Gaussian priors in the symmetric setting (9) is determined by (26); in the general "oblique" ellipsoid case (14), product priors with non-identical components are appropriate. These are typical smoothness priors* for Fourier coefficients. The proof in Pinsker [46] employs also non-Gaussian components, depending on the size of a_j. Bayes-minimax problems in relation to the Pinsker bound are discussed by Heckman and Woodroofe [34], Donoho, MacGibbon and Liu [6], and Donoho and Johnstone [4]. The Pinsker bound thus has a conceptual root both in linear filtering* theory for stochastic processes and in statistical communication* theory. A forerunner of [46] was the result by Ibragimov and Khasminski [36] on the capacity of a Gaussian communication channel under stochastic smoothness restrictions on the signal, expressed in ellipsoid form.

STATISTICAL APPLICATIONS AND FURTHER DEVELOPMENTS

The result of Pinsker [46] for the signal in white noise model (13) or (14) gave rise to a multitude of results in related nonparametric curve estimation problems having a similar structure.

Density estimation and stationary processes. Efromovich and Pinsker [8] treated the case of observed i.i.d. random variables X_j, $j = 1, \ldots, n$ with values in $[0,1]$ having a density f. The result described in (4) was originally obtained for $\mathcal{F}$ being the class of densities in the periodic Sobolev class (23), so that the classical Fourier basis could be used. The proof relies essentially on a kind of uniform local asymptotic normality (LAN) property of the problem, individually for estimation of each Fourier coefficient

$\theta_j(f)$ considered as a statistical functional of f. Similar results were obtained for spectral density estimation for an observed Gaussian stationary sequence, cf. Efromovich and Pinsker [7], Golubev [24], [25].

Nonparametric regression. Consider observations

$$Y_i = f(t_i) + \xi_i,\ i = 1, \ldots, n \tag{28}$$

where ξ_i are i. i. d. $N(0,1)$, $t_i = i/n$ and f is a smooth function on $[0,1]$. Assume again that $f \in \mathcal{F} = W^{m,2}(M)$. Consider a semi scalar product $\langle f, g \rangle_n = n^{-1} \sum_{i=1}^n f(t_i) g(t_i)$ and the associated seminorm $\|f\|_{2,n} = \langle f, f \rangle_n^{1/2}$, and define a minimax risk $R_n(\mathcal{F})$ as in (3) but for a "design loss" $\left\|\hat{f} - f\right\|_{2,n}^2$. Then the asymptotics (4) obtains, cf. [43]. The key for this result is the representation of the model in the ellipsoid form (14), (15). This can be achieved using the Demmler-Reinsch spline basis, which is an orthonormal set of functions $\varphi_{j,n},\ j = 1, \ldots, n$ with respect to $\langle \cdot, \cdot \rangle_n$ and which simultaneously diagonalizes the quadratic form $\left\langle f^{(m)}, f^{(m)} \right\rangle$ (where $\langle \cdot, \cdot \rangle$ denotes scalar product in $L_2(0,1)$). The numbers $a_{j,n} = \left\langle \varphi_{j,n}^{(m)}, \varphi_{j,n}^{(m)} \right\rangle$ represent the coefficients a_j in (15), now depending on n as well. Then the analytic result is required that $a_{j,n} \sim (\pi j)^{2m}$ with appropriate uniformity in n, so that again (22) can be inferred. The optimal estimator of f then is of the linear filtering type in terms of the Demmler-Reinsch spline basis and the Pinsker filter ϕ^* from (24).

Speckman [49] independently found this estimator as minimax linear and gave its risk asymptotics; he used the following setting. Call an estimator $\hat{f}$ of f in (28) linear if it is linear in the n-dimensional data vector Y; then $\hat{f} = AY$ where A is a nonrandom linear operator. The estimator $\hat{f}$ is minimax linear if it minimizes $\sup_{f \in \mathcal{F}} E_{n,f} \left\|\hat{f} - f\right\|_{2,n}^2$ among all linear estimators. In (16) only linear filtering estimates are admitted; it turns out that in the ellipsoid case the minima coincide (cf. also Pilz [45]). Thus another paraphrase of (18) is that the minimax linear estimator is asymptotically minimax among all estimators.

The spectral asymptotics of differential quadratic forms like $\left\langle f^{(m)}, f^{(m)} \right\rangle$ turns out to be crucial, since it governs the behaviour of the ellipsoid coefficients a_j. If spectral values are calculated with respect to $\langle f, f \rangle$ rather than to $\langle f, f \rangle_n$ (which corresponds to continuous observations (13) with parameter space $W_2^m(M)$) then the appropriate basis consists of eigenfunctions of a differential operator, cf. [28], sec. 5.1. The spectral asymptotics is known to be $a_j \sim (\pi j)^{2m}$. The spectral theory for differential operators (cf. Triebel, [50], sect. 5. 6. 2) allows to obtain the Pinsker bound for quite general Sobolev smoothness classes on domains of R^k; for the periodic case on a hypercube domain cf. [42].

Asymptotically Gaussian models. The proof for the cases of density and spectral density estimation ([7], [8]) is based on the asymptotic Gaussianity of those models, in the problem of estimating one individual Fourier coefficient. Inspired by this, Golubev [21] formulated a general local asymptotic normality (LAN) type condition in a function estimation problem, for the validity of the lower bound part of the Pinsker bound. The regression case (28) with nongaussian noise ξ_i in (28) was treated in [28]; for random design regression cf. Efromovich [13].

Analytic functions. The case of m-smooth functions where $a_j \sim (\pi j)^{2m}$ was treated as a standard example here, but another important case in the ellipsoid asymptotics is $a_j \sim \exp(\beta j)$. Then (22) is replaced by

$$R_{L,\epsilon}(\Theta) \sim \left(\epsilon^2 \log \epsilon^{-1}\right) \beta^{-1}.$$

The exponential increase of a_j corresponds to the case of analytic functions; cf. Golubev, Levit, Tsybakov [27]. Ibragimov and Khasminskii [38] obtained an exact risk asymptotics in a case where the functions are even smoother (entire functions of exponential type on the real line) and the rate is ϵ^2, even though the problem is still nonparametric.

Adaptive Estimation. The minimax linear filtering estimate attaining the bound (18) depends on the ellipsoid via the set of coefficients a and M. A significant result of Efromovich and Pinsker [9] is that this attainment is possible even when a and M are not known, provided a varies in some large class of coefficients. The EFROMOVICH-PINSKER ALGORITHM of adaptive estimation (cf. also Efromovich [10]) thus allows to attain the bound (4) for periodic Sobolev classes by an estimator which does not depend on the degree of smoothness m and on the bound M. This represents a considerable advance in adaptive smoothing theory, improving respective rate of convergence results; for further developments and related theory cf. results in [19], [20], [29], [30], [24], [48] and the discussion in [28].

Other constants. Korostelev [41] obtained an analog of the Pinsker bound when the squared L^2-loss $\|\cdot\|_2^2$ is substituted by the sup-norm loss and the Sobolev function class $\tilde{W}_2^m(M)$ is replaced by a Hölder class of smoothness m (a class where f satisfies a condition $|f(x) - f(y)| \leq M\,|x-y|^m$ for all $x, y \in [0,1]$ and given $M > 0$, $m \in (0,1]$). The rate in n then changes to include a logarithmic term and naturally the constant in (22) is another one; this KOROSTELEV CONSTANT represents a further breakthrough and stimulated the search for more constants in nonparametric function estimation. Tsybakov [51] was able to extend the realm of the Pinsker theory to loss functions $w(\|\cdot\|_2)$ where w is monotone and possibly bounded. An analog of the Pinsker bound for nonparametric hypothesis testing was established by Ermakov [14]; cf. also Ingster [39].

Limits of the Pinsker phenomenon. Above it was seen that the case of k-dimensional data (9) and parameter space $\Theta = \{\theta : \sum_{j=1}^{k} \theta_j^2 \leq M\}$ is in some sense the simplest model where the Pinsker phenomenon (18) (asymptotic minimaxity of linear estimators) occurs, as $k, n \to \infty$. Donoho, MacGibbon and Liu [6] set out to investigate more general parameter spaces like $\Theta = \{\theta : \sum_{j=1}^{k} \theta_j^p \leq M\}$ (p-bodies); further results were obtained by Donoho and Johnstone [4]. It was found that (18) occurs only for $p = 2$; linear estimators were found to be asymptotically nonoptimal for $p < 2$, and threshold rules were described as nonlinear alternatives. The limitation of the Pinsker phenomenon to a Hilbertian setting (and thus essentially to L^2-Sobolev classes and related ones) became apparent. However this stimulated the development of nonlinear smoothing methods for other important function classes which cannot be represented as ellipsoids (cf. Donoho and Johnstone [5]).

Further points. Several developments and facets of the theory have not been discussed here; these include applications in deterministic settings ([31], [32], [33]), inverse problems ([15], [16]), design of experiments ([28], [22]), discontinuities at unknown points ([44])

.

Bibliographical remark. The original paper of Pinsker [46] gives the basic idea in a well written first part, but the proof is not easy reading; in addition the English translation as cited is not easy to find. Belitser and Levit [1] present a complete and transparent argument for the basic ellipsoid case in the discrete Gaussian white noise model. Another self-contained but very condensed proof can be found in [43], section 2.

References

[1] Belitser, E., Levit, B., (1995). On minimax filtering on ellipsoids. *Math. Meth. Statist.* **4** 259-273

[2] Belitser, E., Levit, B., Asymptotically minimax nonparametric regression in L_2. *Statistics* 28 105-122

[3] Beran, R. (1996). Stein estimation in high dimensions: a retrospective. In: *Madan Puri Festschrift* (E. Brunner and M. Denker, eds.) 91-110 . VSP Editors, Zeist (Netherlands).

[4] Donoho, D. L. and Johnstone, I, M. (1994). Minimax risk over l_p-balls for l_q-error. *Prob. Theor. Rel. Fields* 99 277-303

[5] Donoho, D. L. and Johnstone, I, M. (1996). Neo-classical minimax problems, thresholding and adaptive function estimation. *Bernoulli* 2 (1) 39-62

[6] Donoho, D. L., Liu, R. and MacGibbon, B. (1990). Minimax risk for hyperrectangles. *Ann. Statist.* **18** 1416-1437.

[7] Efromovich, S. Yu. and Pinsker, M. S. (1981). Estimating a square integrable spectral density based on a sequence of observations. *Problems Inform. Transmission* **17,** 182-196

[8] Efromovich, S. Yu. and Pinsker, M. S. (1982). Estimating a square integrable probability density of a random variable (in Russian). *Problems Inform. Transmission* **18,** No. 3, 19-38

[9] Efromovich, S. and Pinsker, M. S. (1984). An adaptive algorithm of nonparametric filtering. *Automat. Remote Control* **11** 1434-1440

[10] Efromovich, S. (1985). Nonparametric estimation of a density with unkown smoothness. *Theor. Probab. Appl.* **30** 557-568

[11] Efromovich, S. (1989). On sequential nonparametric estimation of a density. *Theor. Probab. Appl.* **34** 228-239

[12] Efromovich, S. (1995). Sequential nonparametric estimation with assigned risk. *Ann. Statist.* **23** 1376-1392

[13] Efromovich, S. (1996). On nonparametric regression for iid observations in a general setting. *Ann. Statist.* **24** 1126-1144

[14] Ermakov, M. S. (1990). Minimax detection of a signal in Gaussian white noise. *Theory Probab. Appl.* **35** 667-679

[15] Ermakov, M. (1990). On optimal solutions of the deconvolution problem. *Inverse Problems* **6** 863-872

[16] Ermakov, M. (1992). Minimax estimation in a deconvolution problem. *J. Phys. A: Math. Gen.* 1273-1282

[17] Golubev, G. K. (1982). On minimax filtering of functions in L_2 (in Russian). *Problems Inform. Transmission* **18,** No. 4, 67-75

[18] Golubev, G. K. (1984). On minimax estimation of regression (in Russian). *Problems Inform. Transmission* **20,** No. 1, 56-64

[19] Golubev, G. K. (1987). Adaptive asymptotically minimax estimates of smooth signals. (in Russian). *Problems Inform. Transmission* **23,** No. 1, 57-67

[20] Golubev, G. K. (1990). Quasilinear estimates for signals in L_2. *Problems Inform. Transmission* **26**, 1, 15-20

[21] Golubev, G. K. (1991). Local asymptotic normality in problems of nonparametric estimation of functions, and lower bounds for quadratic risk. *Theory Probab. Appl.* **36,** 1, 152-157

[22] Golubev, G. K. (1992). Sequential design of an experiment in the nonparametric estimation of smooth regression functions. *Problems Inform. Transmission* **28**, 3, 265-268

[23] Golubev, G. K. (1992). Asymptotically minimax estimation of a regression function in an additive model. *Problems Inform. Transmission* **28**, 2, 101-112

[24] Golubev, G. K. (1992). Quasilinear filtering of Gaussian stationary sequences. *Theory Probab. Appl.* **37**, 4, 793-800

[25] Golubev, G. K. (1993). Nonparametric estimation of smooth spectral densities of Gaussian stationary sequences. *Theory Probab. Appl.* **38** 4 630-639

[26] Golubev, G. K., Levit, B. (1995). On the second order minimax estimation of distribution functions. *Math. Methods of Statistics,* to appear

[27] Golubev, G. K., Levit, B., Tsybakov, A. B. (1996). Asymptotically efficient estimation of analytic functions in Gaussian white noise. *Bernoulli* **2** 167-181

[28] Golubev, G. K. and Nussbaum, M. (1990). A risk bound in Sobolev class regression. *Ann. Statist.* **18** 758-778

[29] Golubev, G. K. and Nussbaum, M. (1990). Nonparametric estimation of a regression function in L_2. *Problems Inform. Transmission* **26,** 3, 213-255

[30] Golubev, G. K. and Nussbaum, M. (1992). Adaptive spline estimates in a nonparametric regression model. *Theory Probab. Appl.* **37**, 3, 521-529

[31] Golubev, G. K. and Pinsker, M. S. (1983). Minimax extrapolation of sequences. *Problems Inform. Transmission* **19**, 4, 275-283

[32] Golubev, G. K. and Pinsker, M. S. (1984). Minimax extrapolation of functions. *Problems Inform. Transmission* **20**, 2, 27-43

[33] Golubev, G. K. and Pinsker, M. S. (1985). Extremal properties of minimax estimation of sequences. *Problems Inform. Transmission* **21**, 3, 36-52

[34] Heckman, N. E. and Woodroofe, M. (1991). Minimax Bayes estimation in nonparametric regression. *Ann. Statist.* **19** 2003-2014

[35] Hasminski, R., and Ibragimov, I. (1990). On density estimation in the view of Kolmogorov's ideas in approximation theory. *Ann. Statist.* **18** 999-1010

[36] Ibragimov, I. A. and Khasminskii, R. Z. (1979). Channel capacity under smoothness restrictions on the signal. *Problems Inform. Transmission* **15**, 3, 18-26

[37] Ibragimov, I. A. and Khasminskii, R. Z. (1981). *Statistical Estimation: Asymptotic Theory.* Springer, New York.

[38] Ibragimov, I. A. and Khasminskii, R. Z. (1982). On density estimation within a class of entire functions. *Theory Probab. Appl.* **27** 514-524

[39] Ingster, Yu. I. (1993). Asymptotically minimax hypothesis testing for nonparametric alternatives. *Math. Meth. Statist.,* Part I: **2** 2 85-114, Part II: **2** 3 171-189, Part III: **2** 4 249-268

[40] Kazbaras, A. (1986). An adaptive kernel type estimator for a square integrable distribution density. *Lithuanian Math. Journal* **26** 318-324

[41] Korostelev, A. P. (1993). Exact asymptotically minimax estimator for nonparametric regression in uniform norm. *Theory Probab. Appl.* **38**, no. 4, 775–782

[42] Nussbaum, M. (1983). Optimal filtering of a function of many variables in Gaussian white noise. *Probl. Information Transmission* **19** (2) 23-29 (1983)

[43] Nussbaum, M. (1985). Spline smoothing in regression models and asymptotic efficiency in L_2. *Ann. Statist.* **13** 984-997

[44] Oudshoorn, K. (1996). Minimax estimation of a regression function with jumps: attaining the optimal constant. To appear, *Bernoulli.*

[45] Pilz, J. (1986). Minimax linear regression estimation with symmetric parameter restrictions. *J. Statist. Plann. Inference* **13** 297-318

[46] Pinsker, M. S. (1980). Optimal filtering of square integrable signals in Gaussian white noise (in Russian). *Problemy Peredachi Informatsii* 16 (2) 52-68, English translation *Problems Inform. Transmission* (1980) 120-133

[47] Rudzkis, R. (1985). On an estimate of the spectral density. *Lithuanian Math. Journal* **25** 273-280

[48] Locally minimax efficiency of nonparametric estimates of square-integrable densities. *Lithuanian Math. Journal* **33** 56-75

[49] Speckman, P. (1985). Spline smoothing and optimal rates of convergence in nonparametric regression. *Ann. Statist.* 13 970-983

[50] Triebel, H. (1978). *Interpolation theory, Function Spaces, Differential Operators.* North-Holland, Amsterdam.

[51] Tsybakov, A. B. (1997). Asymptotically efficient signal estimation in L_2 under general loss functions. *Problems of Information Transmission* ***33*** *(1)* .